XML, RDF, RSS, and XSLT
A Mixture of Technologies

Norman Walsh

XML.Gov Working Group
10 September, 2003

http://www.sun.com/

Sun microsystems
We make the net work.

Version 1.0

About Norm

- XML Standards Engineer at Sun Microsystems, Inc.
- Elected member of the W3C Technical Architecture Group (TAG).
- Member of the XML Core, XSL, and other WGs at W3C.
- Chair of the DocBook Technical Committee at OASIS.
- Member of the RELAX NG and Entity Resolution TCs.

About this Presentation

- For the first version of the Architecture of the World Wide Web, the TAG is focussed mostly on Identity, Representation, and Interaction.
- Waiting in the wings are the Semantic Web and Web Services.
- Discussing WebArch in detail raises a lot of interesting questions.
- I built `norman.walsh.name` to explore these questions.
- That site attracted the interest of at least some of the folks involved in the XML.Gov WG.

Goals

- Describe a set of related technologies.
- Explore how they can be used to complement each other.
- Provide enough detail so that you can see how they might apply to the problems you're facing.
- Leave enough time for discussion.

Web Site as Information System Microcosm

- Content (essays, images, applications, etc.)
- Metadata (content about the content)
- Content access (navigation, search, etc.)
- Content delivery (multiple representations, content negotiation, etc.)
- Notification (what's new, what's changed?)
- Update (revising content, adding new content)

Content = XML

- XML is tree structured.
- There's a more-or-less 1:1 mapping between serialized XML documents and the trees they represent.
- XML documents “stand alone”. Given an XML document, there's no general concept of mechanically “merging” it with another document.

Metadata = RDF

- RDF is graph structured.
- There are several ways to serialize an RDF graph.
- Given an RDF graph, it's straightforward to add new information to that graph from another RDF graph.

Notification = RSS

- RSS is the site summary or syndication format.
- There are several versions and work is actively progressing on a “next generation” version.
- It's a mostly grass-roots development process.
- There's no “official” standards organization involved.
- Use it anyway.

Access/Delivery/Update

- Access = Metadata as content.
- Delivery = Sending what the user wants (XML, HTML, PDF, etc.). (There are some interesting architectural and technical issues here.).
- Update = Adding content, providing for feedback.

RDF Concepts

- RDF stands for Resource Description Framework.
- Developed by the W3C as part of the Semantic Web Activity.
- A framework for representing metadata statements.
- A metadata statements is a simple statement about an object: nothing more, nothing less.
- Every statement has three parts: a subject, a predicate, and a value for the predicate (an object).
- RDF is really just a collection of simple statements (often called “triples”).

RDF Statements

Some RDF statements (informally):

NormanWalsh has blue eyes
IRS type GovtAgency
ThisDocument creator NormanWalsh
Meeting agenda http://xml.gov/agenda/20030910.htm
hisDocument date "2003-09-10"
BWI lat "39.1753611"
BWI long "-76.6683333"
BWI label "BWI"
BWI type Airport
BWI weather http://www.weather.com/weather/local/21240

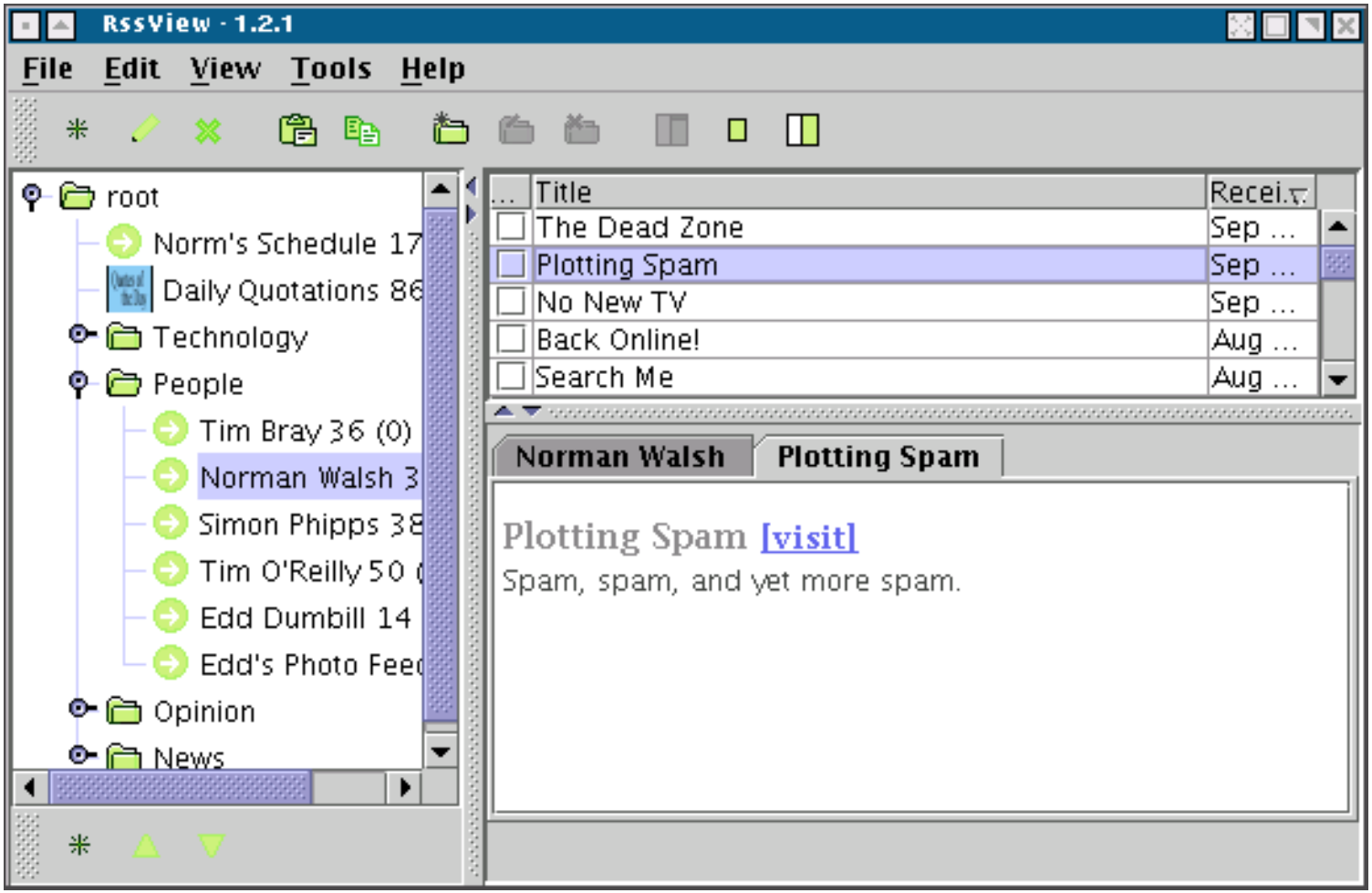
Why RDF?

- Extends easily: new subjects, new predicates.
- Combines easily: building aggregate metadata is easy.
- Software can apply logic to statements (concluding that the author of ThisDocument has blue eyes, for example).
- There are tools (cwm, Jena) that make it possible to solve interesting problems with RDF.

RSS

- An XML (and sometimes RDF) vocabulary.
- Popular format for publishing summaries (What's new? What's available?).
- End-user tools for aggregation.
- Business models based on aggregation.

RSS Viewer


RssView - 1.2.1
File Edit View Tools Help
root
Norm's Schedule 17
Daily Quotations 86
Technology
People
Tim Bray 36 (0)
Norman Walsh 3
Simon Phipps 38
Tim O'Reilly 50
Edd Dumbill 14
Edd's Photo Feed
Opinion
News
Title
Recei...
The Dead Zone
Sep ...
Plotting Spam
Sep ...
No New TV
Sep ...
Back Online!
Aug ...
Search Me
Aug ...
Norman Walsh
Plotting Spam
Plotting Spam [visit]
Spam, spam, and yet more spam.

Other RSS Applications

- Web site updates.
- Magazine and journal tables of contents.
- Updates and notifications of all kinds.
- I keep my daily schedule in an RSS feed, with additional metadata for task categories and priorities.

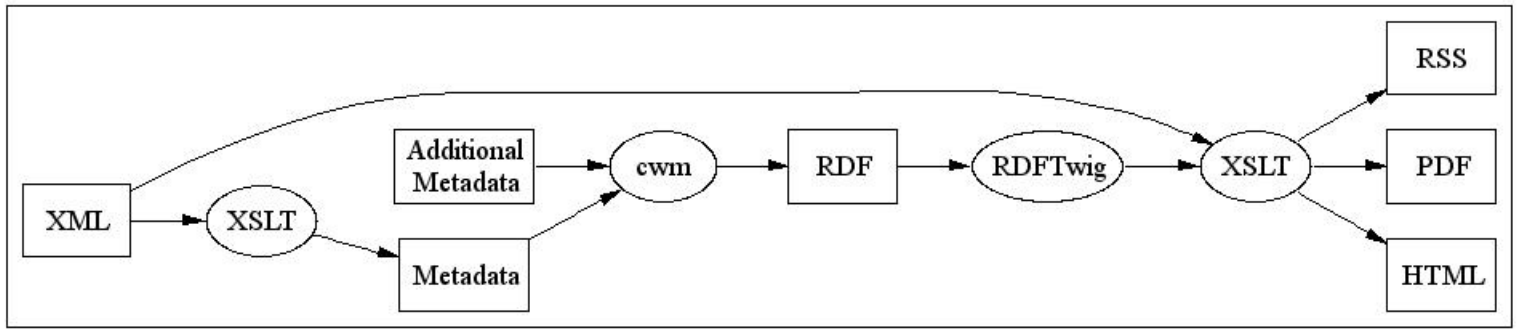
XSLT

- XML Transformations (tree-to-tree transforms; it's really a full programming language, albeit a slighty odd one).
- Widely used to produce online and print rendering of XML data sources.
- XML to HTML (web publishing).
- XML to FO to PDF (print publishing).
- XML to RSS (summaries).
- XML to RDF (metadata extraction).

XSLT and RDF: Oil and Water?

- RDF is a graph.
- XSLT is a tree transformation language.
- Graphs are not trees. Graphs have many serializations.
- There are a number of approaches to this problem.
- RDFTwig is my own.

Putting the Pieces Together



Information Flow

- Most metadata comes from XML.
- Additional metadata maintained by hand.
- cwm and Jena used to manage the RDF.
- XSLT (and RDF TWig) used for transformations.

Successes

- RDF lets me easily aggregate metadata.
- Driving all the presentation from the metadata lets me ignore the fact that the data is aggregated.
- Storing metadata separately allows me to modify the entire site by changing only a single file.

Failures

- There's more complexity in the build system.
- There may be scaling issues.

Conclusions

- XML is perfect for storing structured information.
- Explicitly storing metadata separately has advantages.
- RDF is a useful technology for managing metadata.
- RSS keeps everyone (who cares) up-to-speed.
- XSLT can be coerced into working with RDF, making it applicable to both data and metadata.

References

- RDF, http://www.w3.org/RDF/.
- RSS, http://www.purl.org/rss/1.0/, http://www.intertwingly.net/wiki/pie/RSS.
- XSLT, http://www.w3.org/TR/xslt.
- cwm, http://www.w3.org/2000/10/swap/doc/cwm.html.
- Jena, http://www.hpl.hp.com/semweb/.
- rssview, http://rssview.sourceforge.net/.
- RDFTwig, http://rdftwig.sourceforge.net/.

RDF Technicalities

- Statements are only useful if we can agree on what the subjects, predicates, and values mean.
- Use URIs to identify subjects and predicates. (Globally unique names, essentially the same reason we use QNames in XML.)
- Values can be simple literals (1, 3.14, "Norman Walsh") or URIs, meaning that the values themselves can be the subject of other statements.
- Can be serialized in XML, in a text format called N3, and other ways.

RDF Statements Revisited

The RDF statements again (in N3):

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ex: <http://example.com/rdf/vocabulary#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix wn: <http://xmlns.com/wordnet/1.6/>

<http://norman.walsh.name/knows/who#norman-walsh> ex:has "blue eyes" .
<http://www.irs.gov> rdf:type ex:GovtAgency .
<> dc:creator <http://norman.walsh.name/knows/who#norman-walsh> .
<http://xml.gov/> ex:agenda <http://xml.gov/agenda/20030910.htm> .
<> dc:date "2003-09-10" .
```

RDF Statements Revisited (Continued)

```
ex:BWI geo:lat "39.1753611" .
ex:BWI geo:long "-76.6683333" .
ex:BWI rdfs:label "BWI" .
ex:BWI rdf:type wn:Airport .
ex:BWI ex:weather <http://www.weather.com/weather/local/21240> .
```